BAYESIAN PREDICTIVE PROCESS MODELS FOR HISTORICAL PRECIPITATION DATA OF ALASKA AND SOUTHWESTERN CANADA

By

Peter Vanney

RECOMMENDED:

Dr. Ronald Barry

Dr. Scott Goddard

Dr. Margaret Short
Advisory Committee Chair

Dr. John Rhodes, Chair
Department of Mathematics and Statistics

BAYESIAN PREDICTIVE PROCESS MODELS FOR HISTORICAL PRECIPITATION DATA OF ALASKA AND SOUTHWESTERN CANADA

A

PROJECT

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Peter Vanney, B.A.

Fairbanks, AK

May 2016

Abstract

In this paper we apply hierarchical Bayesian predictive process models to historical precipitation data using the **spBayes** R package. Classical and hierarchical Bayesian techniques for spatial analysis and modeling require large matrix inversions and decompositions, which can take prohibitive amounts of time to run (n observations take time on the order of n3). Bayesian predictive process models have the same spatial framework as hierarchical Bayesian models but fit a subset of points (called knots) to the sample which allows for large scale dimension reduction and results in much smaller matrix inversions and faster computing times. These computationally less expensive models allow average desktop computers to analyze spatially related datasets in excess of 20,000 observations in an acceptable amount of time.

1 Introduction

The availability of large spatial datasets has increased drastically over the last 15 years thanks to an increased use of geographical information systems and global positioning systems in science. Books by Cressie (1993) and Banerjee, Carlin and Gelfand (2004) analyze and model geostatistical spatial data (among other types of spatial data) by defining the spatial relationship over the region of interest, but many of the traditional techniques they employ are limited by n, the sample size of the dataset. This "big n problem" arises from matrix inversions and decompositions requiring time on the order of n^3 to complete. However, articles by Higdon (2002), Banerjee et al. (2008), Finley et al. (2008), Banerjee and Fuentes (2012), Eidsvik et al. (2012) and Finley, Banerjee and Gelfand (2015) have focused on dimension reduction techniques to lower the computing time of large sample models. We make use of predictive process models, pioneered by Banerjee which employ a fixed set of knots of size $m \ll n$ and only require computations $O(m^3)$.

Stein (2004 and 2008), Cressie and early Higdon (2002) specify various frequentist models for geostatistical spatial data. They estimate model parameters using maximum likelihood and restricted maximum likelihood techniques. More recently Higdon, Cressie, Banerjee and others use hierarchical Bayesian models, typically fitted using Markov Chain Monte Carlo methods. Hierarchical Bayesian techniques allow model parameters to be defined as random variables dependent on hyperparameters which allows the models to be more flexible.

The large climatological maps created by SNAP (Scenarios Network for Alaska + Arctic Planning), using PRISM (Parameter-elevation Regressions on Independent Slopes Model) motivated the use of hierarchical Bayesian predictive process models in this paper, specifically the dataset containing precipitation from August 2013 across Alaska and Southwestern Canada. The map contains over 3.8 million gridded locations, 1.4 million of which contain numerical observations.

The two main objectives of this paper are to model part of a specific SNAP map and

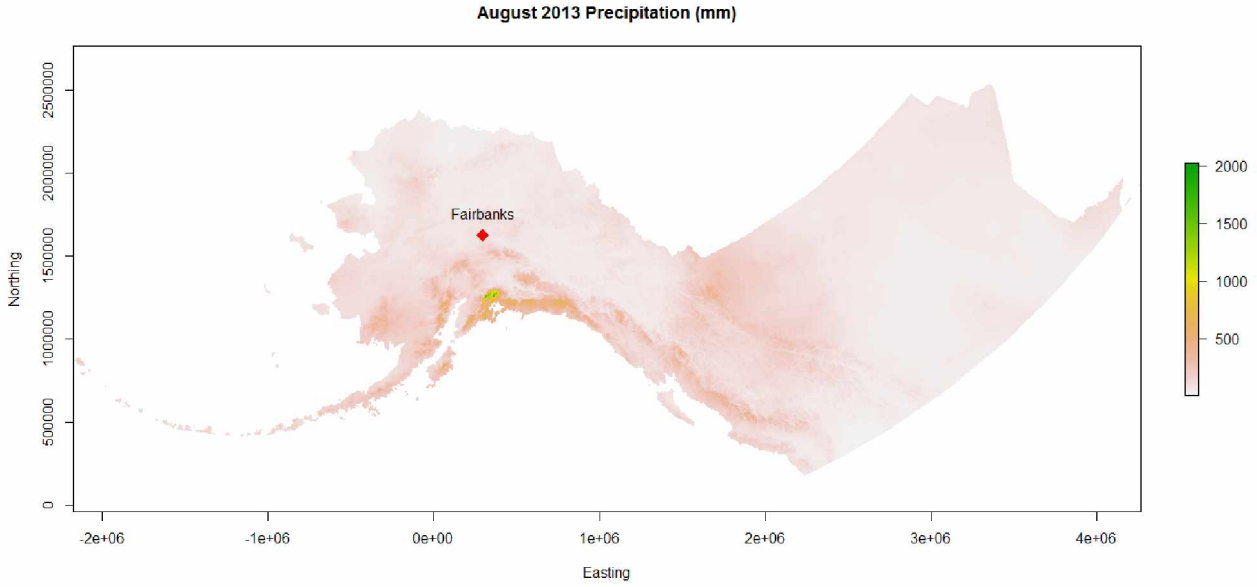


Figure 1: SNAP map of total precipitation across Alaska and Southwestern Canada in August 2013

to explore predictive process models in general. Besides the general issue of estimating a map from a large number of irregularly spaced and randomly sampled points, estimates of overall variability of the map were of interest. Using a predictive process model also allowed us to observe how the variability and model fitting time changed as the number of knots or the sample size increased.

This paper is organized as follows: we explain the dataset in Section 2; we provide an overview of geostatistical analysis and a description of the model in Section 3; we present the results of different model runs in Sections 4 and 5; and finally we end the paper with a discussion in Section 6.

2 Data

2.1 PRISM and SNAP

The PRISM Climate Group at Oregon State University has been creating large scale precipitation and temperature maps for the contiguous 48 states since the early 1990s when Christopher Daly first developed the PRISM climate mapping system. The precipitation system combines data from weather stations with topographic maps in such a way that the predicted rainfall is primarily dependent on elevation and the slope orientation of the terrain (N-facing, SE-facing, etc.). Each orientation category is calculated independently, and the estimates are smoothed across neighboring features, taking into consideration distance to coastlines, known temperature inversions, and other measures of terrain complexity. The modeling system is known as PRISM (Parameter-elevation Regressions on Independent Slopes Model). Shortly after the first version was created, the Natural Resources Conservation Service of the United States Department of Agriculture partnered with Oregon State University to fund the PRISM Climate Group. (Daly et al. (2008); Daly and Bryant (2013)).

SNAP (Scenarios Network for Alaska + Arctic Planning) is located at the University of Alaska Fairbanks. It aims to develop and communicate plausible climate change scenarios through collaboration and modeling. Figure 1 is an example of the PRISM model applied to Alaska and Southwestern Canada by SNAP. The map shows the monthly precipitation total in millimeters for the region on a 2km$\times$2km grid; the values on the grid are known as pixels. The map contains 3.8 million pixels, although 2.4 million are missing values because they are outside the boundary, leaving 1.4 million observations. (*SNAP* (2015)).

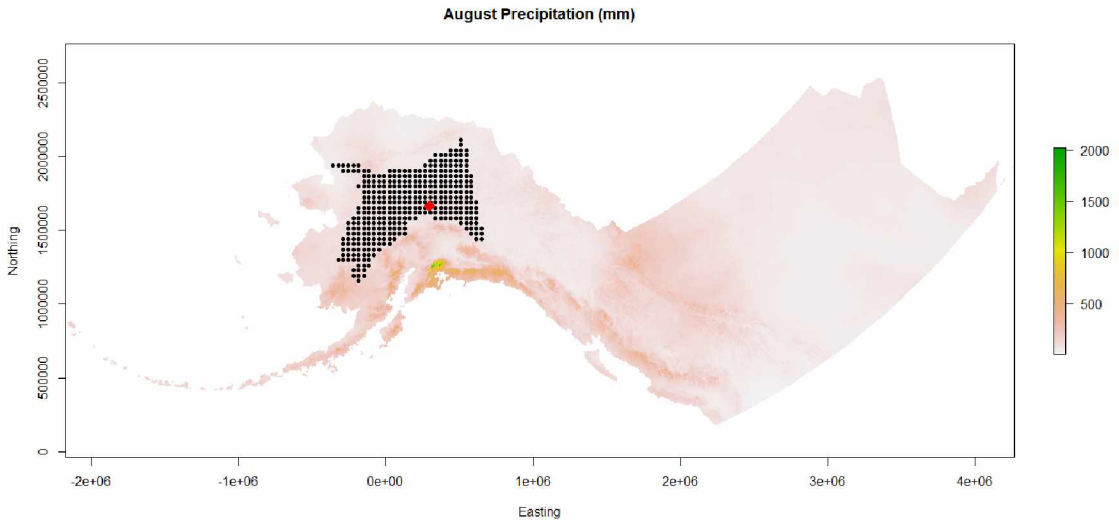


Figure 2: Original dataset with the Northern Boreal Forest shaded

2.2 Northern Boreal Forest

Within the map in Figure 1 there are seven landscape conservation cooperatives (LCCs): Northwest Boreal, Plains and Prairie Potholes, Aleutian and Bering Sea Islands, Western Alaska, North Pacific, Great Northern, and Arctic. LCCs are federally designated areas that transcend political boundaries and rely on partnerships between public and private organizations for effective conservation. SNAP is one of the collaborators that supports LCCs, and they separate the Northwest Boreal LCC into a northern and southern region. For the purposes of this paper, we restrict the analysis to the northern section of the Northwest Boreal LCC in Alaska, and we simply refer to it as the Northern Boreal Forest. Figures 2 and 3 show the location and precipitation of the Northern Boreal Forest, respectively.

The Northern Boreal map contains only 114,000 observations on 268,000 total pixels; however, it has many of the same features as the larger map such as an irregular border and non-normally distributed data. Figure 4 shows the distribution of observations over the region. The upper panel illustrates the spatial clustering over the forest: the dark,

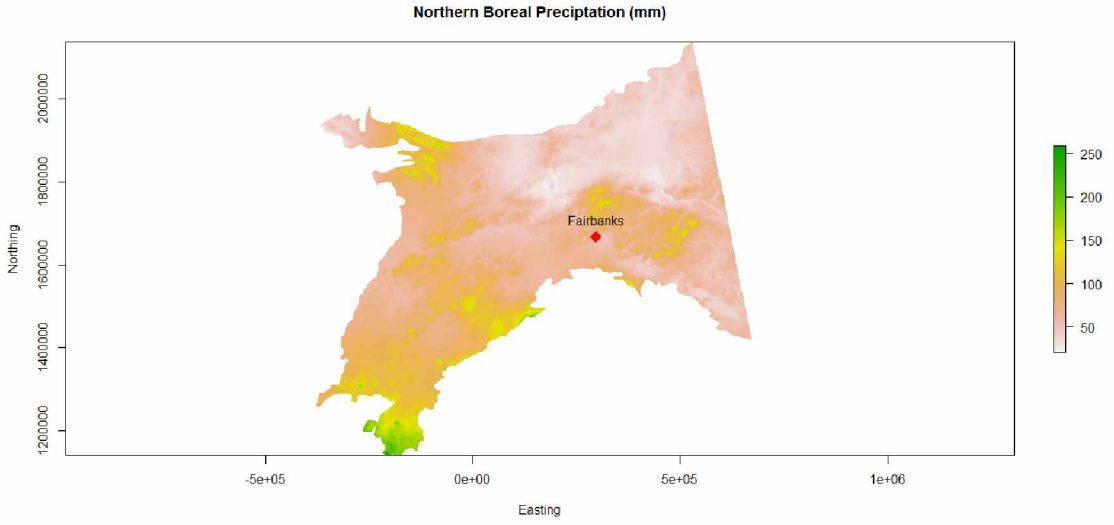
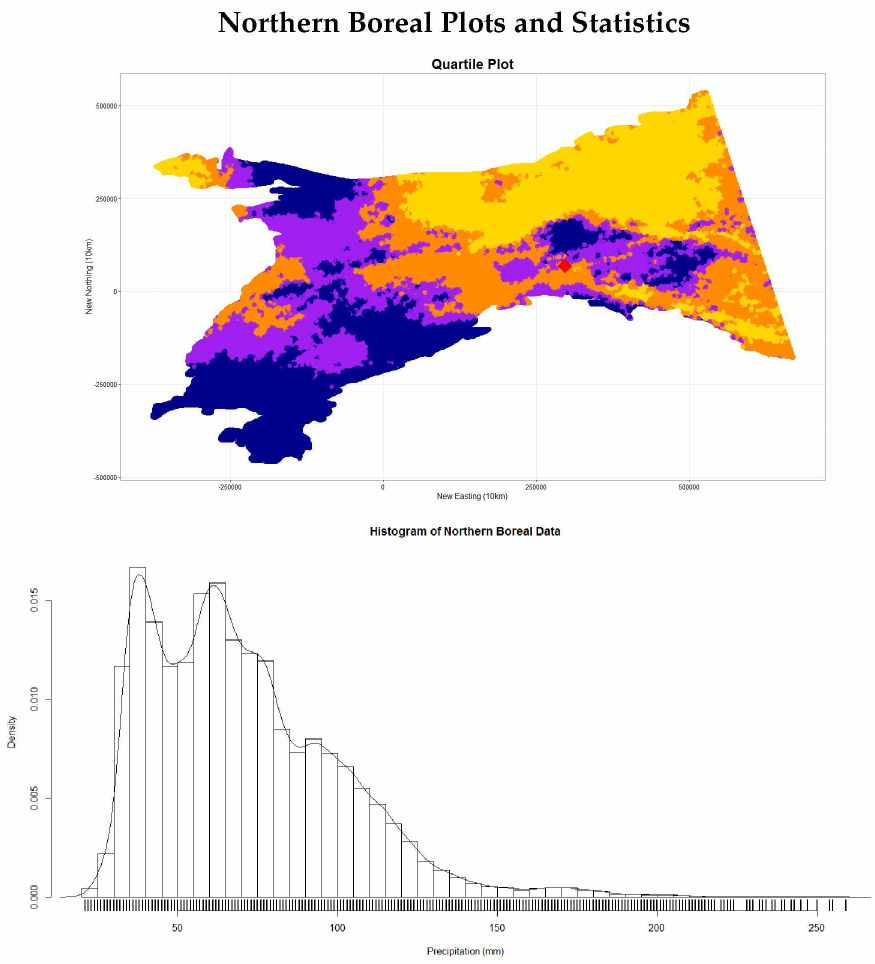


Figure 3: The Northern Boreal Forest data

southwestern corner of the forest receives the most rain while the lighter northcentral region receives the least. We expect that any predictions would follow a similar trend. The lower panel shows the non-normal, right-skewedness of the data, which our model needs to be able to accommodate.



Minimum	Maximum	Median	Mean	Std. Deviation
21	259	66	71.05	29.74

Figure 4: The first plot shows the values grouped by quartile, and the histogram shows the density of the observations.

3 Model

3.1 Geostatistical Data Analysis

The customary model for geostatistical analysis assumes that for locations $\mathbf{s} \in \mathcal{D}$, a response variable $Y(\mathbf{s})$ and predictors $\mathbf{x}(\mathbf{s})$ are related through the model

$$Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

where $\mathbf{x}(\mathbf{s})$ is a $p \times 1$ vector of location-specific covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients, $\varepsilon(\mathbf{s})$ is a nugget effect, and we assume $\varepsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$ for every location $\mathbf{s}$ and $w(\mathbf{s})$ is a mean 0 Gaussian process with covariance function $c(\boldsymbol{\theta}) = c(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}) = \text{Cov}(w(\mathbf{s}), w(\mathbf{s}'))$. Given a sample of geostatistical data $\mathbf{Y} = [Y(s_1), \ldots, Y(s_n)]^T$, $c(\boldsymbol{\theta})$ induces an $n \times n$ covariance matrix $C(\boldsymbol{\theta})$, and by applying this model we find $\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}, \tau^2 \sim MVN(\mathbf{X}\boldsymbol{\beta}, \Sigma_Y)$, where $\Sigma_Y = C(\boldsymbol{\theta}) + \tau^2\mathbf{I}$. The likelihood function is then given by

$$L(\tau^2, \boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{Y}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_Y|}} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \Sigma_Y^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}. \tag{1}$$

Using this form we could estimate the parameters with restricted maximum likelihood or maximum likelihood methods, or we could specify priors and sample from the posterior distribution using Markov Chain Monte Carlo methods. However, both frequentist and Bayesian methods require inversions or decompositions of the $n \times n$ matrix Σ_Y at each iteration taking time $O(n^3)$. This "big n problem" leads us to explore the predictive process model as an alternative.

3.2 Predictive Process Model

The predictive process model has the same general form as the customary model, but we define a new spatial process $\tilde{w}(\mathbf{s})$ to replace $w(\mathbf{s})$ in the model. First, define a set of knots $S^* = \{s_1^*, s_2^*, \ldots, s_m^*\}$ at locations not necessarily within $\mathcal{D}$, such that $m \ll n$. These locations might coincide with some or all of the observation locations $\{s_1, s_2, \ldots, s_n\}$, but they need not. Let $\boldsymbol{w}^* = (w(s_1^*), w(s_2^*), \ldots, w(s_m^*))^T$ be an $m \times 1$ vector of spatial random effects, so that $\boldsymbol{w}^* \sim MVN(\mathbf{0}, C^*)$ where $(C^*)_{ij} = c((s_i^*, s_j^*); \boldsymbol{\theta}) = \text{cov}(w(s_i^*), w(s_j^*))$. Then $\boldsymbol{w}^*$ has probability density function

$$p(\boldsymbol{w}^*|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^m |C^*|}} \exp\left\{-\frac{1}{2}(\boldsymbol{w}^*)^T (C^*)^{-1} (\boldsymbol{w}^*)\right\}.$$

Now we define a new spatial process $\tilde{w}(s)$. For any arbitrary spatial location s, define

$$\tilde{w}(s) = c(s;\boldsymbol{\theta})^T (C^*)^{-1} \boldsymbol{w}^*$$

where $c(s;\boldsymbol{\theta}) = [c(s,s_1^*), c(s,s_2^*), \ldots, c(s,s_m^*)]^T$. Note that $\tilde{w}(s)$ depends on s solely through $c(s;\boldsymbol{\theta})^T$ because $(C^*)^{-1}$ and $\boldsymbol{w}^*$ depend only on S^*.

We note that $\tilde{w}(\mathbf{s})$ is a mean 0 Gaussian process with covariance function $c(s;\boldsymbol{\theta})^T (C^*)^{-1} c(s';\boldsymbol{\theta})$, as follows:

$$\begin{aligned}
\text{cov}[\tilde{w}(s), \tilde{w}(s')] &= \text{cov}[c(s;\boldsymbol{\theta})^T (C^*)^{-1} \boldsymbol{w}^*, c(s';\boldsymbol{\theta})^T (C^*)^{-1} \boldsymbol{w}^*] \\
&= c(s;\boldsymbol{\theta})^T (C^*)^{-1} \text{Cov}[\boldsymbol{w}^*][c(s';\boldsymbol{\theta})^T (C^*)^{-1}]^T \\
&= c(s;\boldsymbol{\theta})^T (C^*)^{-1} C^* (C^*)^{-1} c(s';\boldsymbol{\theta}) \\
&= c(s;\boldsymbol{\theta})^T (C^*)^{-1} c(s';\boldsymbol{\theta})
\end{aligned}$$

As before, given a sample of geostatistical data $\mathbf{Y} = [Y(s_1), \ldots, Y(s_n)]^T$, and using the predictive process model in which $\tilde{w}(\mathbf{s})$ has replaced $w(\mathbf{s})$, we have the following model

$$Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + \tilde{w}(\mathbf{s}) + \varepsilon(\mathbf{s}).$$

In equation (1), the spatial effect term $w(\mathbf{s})$ has been integrated out. However, if we condition on $\tilde{w}(\mathbf{s})$ as a parameter, then we get

$$\begin{aligned}
Y_i|\tau^2, \boldsymbol{\theta}, \boldsymbol{\beta}, \tilde{w}(s_i) &\overset{ind}{\sim} N\left(\mathbf{x}(s_i)^T\boldsymbol{\beta} + \tilde{w}(s_i), \tau^2\right) \\
&\overset{ind}{\sim} N\left(\mathbf{x}(s_i)^T\boldsymbol{\beta} + c(s_i;\boldsymbol{\theta})^T (C^*)^{-1}\boldsymbol{w}^*, \tau^2\right)
\end{aligned}$$

which only depends on $\tilde{w}$ through $\boldsymbol{w}^*$ because $\tilde{w}$ is a deterministic function of $\boldsymbol{w}^*$. Thus, maximum likelihood or MCMC methods require time $O(m^3)$ for operations on C^*

instead of $O(n^3)$ for operations on Σ_Y to determine parameters of the following likelihood

$$L(\tau^2, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{w}^*; \mathbf{Y}) = \frac{1}{\sqrt{(2\pi)^n(\tau^2)^n}} \exp\left\{-\frac{1}{2\tau^2}\left|\left|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \tilde{w}(\mathbf{s})\right|\right|^2\right\}. \tag{2}$$

In the predictive process model and the traditional model, $\boldsymbol{\theta}$ is usually a vector consisting of two or more parameters that determine the covariance between two points based on the distance between them. For example, for an exponential covariance model, $\boldsymbol{\theta} = (\sigma^2, \phi)$ and $c(s_i, s_j; \boldsymbol{\theta}) = \sigma^2 e^{-\phi||s_i - s_j||}$.

Now we can finally write the full posterior distribution used for Bayesian inference.

$$\begin{aligned} p(\tau^2, \sigma^2, \phi, \boldsymbol{\beta}, \boldsymbol{w}^*|\mathbf{Y}) &\propto L(\tau^2, \sigma^2, \phi, \boldsymbol{\beta}, \boldsymbol{w}^*; \mathbf{Y}) * p(\boldsymbol{w}^*|\sigma^2, \phi) * \pi(\tau^2, \sigma^2, \phi, \boldsymbol{\beta}) \\ &\propto L(\tau^2, \sigma^2, \phi, \boldsymbol{\beta}, \boldsymbol{w}^*; \mathbf{Y}) * p(\boldsymbol{w}^*|\sigma^2, \phi) * \pi(\tau^2) * \pi(\sigma^2) * \pi(\phi) * \pi(\boldsymbol{\beta}) \end{aligned}$$

It is natural to wonder if $\boldsymbol{\theta} = (\sigma^2, \phi)$ in the exponential model is equivalent or has an equivalent interpretation in the traditional and predictive process models. Banerjee et al. (2008) focus on comparisons between $C(\boldsymbol{\theta})$ and C^* to conclude that $\boldsymbol{\theta}$ is roughly equivalent under certain conditions. Given locations s and s', where $s' = s + h$ for spatial lag h,

$$\begin{aligned} \text{Cov}(w(s), w(s')) &\approx \text{Cov}(\tilde{w}(s), \tilde{w}(s')) \\ c(s, s'; \boldsymbol{\theta}) &\approx c(s; \boldsymbol{\theta})^T (C^*)^{-1} c(s'; \boldsymbol{\theta}) \end{aligned}$$

when $||h||$ is large. Banerjee et al. note that what matters most is the size of the range compared to the spacing of knots, and that having knots spaced further apart than $||h||$ results in larger differences between the two covariance functions for small $||h||$. Hence, for fine scale covariance structure modeling, at least some knots should be positioned close together.

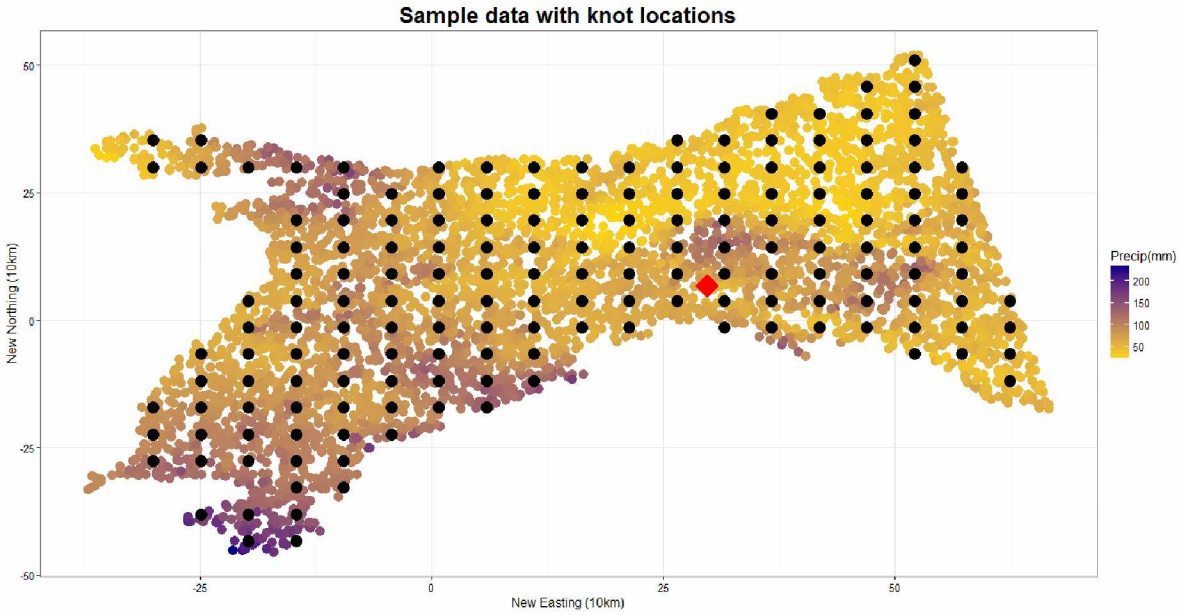


Figure 5: Northern Boreal subsample of size $n = 5{,}000$ with $m = 170$ knots

3.3 Picking knots

Because the motivation for this paper was recreating maps, we simply used a regularly spaced grid of knots. It was important that they span the whole region; since we were initially less concerned with the spatial parameter estimates, we did not add any infill to capture fine-scale covariance structure. Also, since the data were generated and smoothed using PRISM, it is reasonable to assume there would be no additional variability between very close observations. Variability between extremely close coordinates is modeled by τ^2 and is known as the nugget; this error term accounts for any other unexplained variability as well. The number of knots is mostly limited by computation time and relative performance of the models. For most of the models we used 170 knots, but overall map fit improved as the number of knots increased.

Although the predictive process model may make use of knots outside of the sampled region, determining knot values outside of the Northern Boreal Forest seemed a wasted effort. By first selecting knots on a larger grid and then removing knot locations outside of the region, we were able to increase the knot density within the Northern Boreal Forest

without computing unnecessary values. Figure 5 shows the knot coverage for $m = 170$ knots and a subsample size of $n = 5{,}000$.

3.4 spBayes and spLM

The predictive process model was implemented using the **spBayes** R package and the function spLM by specifying knot coordinates, as well as starting values, tuning parameters, the type of covariance function, covariates, and prior distributions.

All of the models we compared used an exponential covariance function, but we fit models using intercept-only main effects and main effects that included an intercept, Easting, Northing and their interaction as regressors. For priors we used

$$\beta_i \overset{ind}{\sim} N(0, \sigma^2 = 1{,}000)$$
$$\phi \sim Uniform(.01, 20)$$
$$\sigma^2 \sim InverseGamma(shape = 3, scale = 2).$$

Note that in this parameterization of the exponential covariance function, ϕ is a rate parameter instead of the customary scale parameter; thus, it has units of $distance^{-1}$. The starting values had little effect on model convergence. However, the tuning parameters of β, σ^2 and ϕ proved to be more difficult and needed to be set low because the parameters are not independent. The trace plots in Figure 10 in Appendix A.1 have matching spikes which illustrate this dependence, and the scatterplot in Figure 11 illustrates the dependence between σ^2 and ϕ.

As previously mentioned, because the data were created using PRISM, there is no need for a nugget term, τ^2. Models without τ^2 can be implemented using spLM by omitting the prior distribution and starting value for τ^2 when specifying a model. However, without a nugget the likelihood function of the predictive process model has zero variance, so spLM uses a modified predictive process model which sets τ^2 to a small, nonzero constant and

adds jitter to any knot coordinates that overlap with observations.

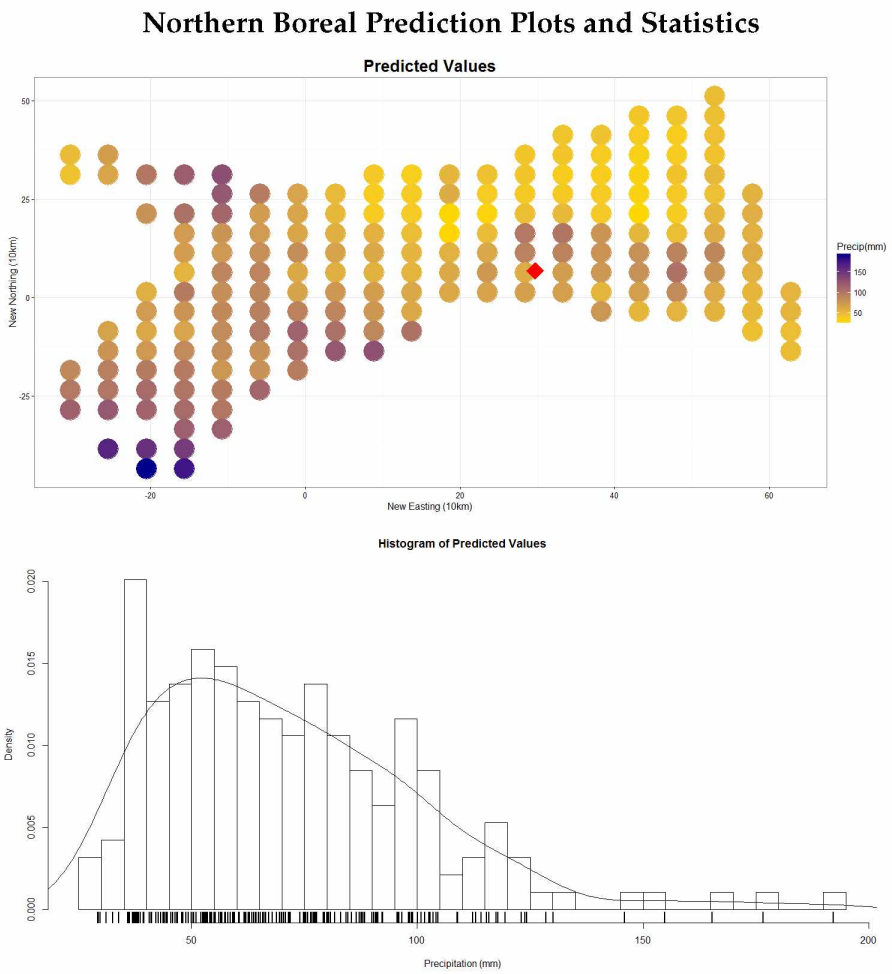
While performing MCMC, `spLM` only saves the values of β and $\boldsymbol{\theta}$ at each iteration. To obtain the knot values $w^*(\mathbf{s})$ one must use the function `spRecover` which samples from the posterior distributions of the specified parameters. Similarly, the function `spPredict` is used to make predictions at unsampled locations using a conditional multivariate normal model based on the observed data and samples from the joint posterior distribution of the parameters. In order to compare results and predictions more easily, we predicted values at 189 evenly spaced locations for all models.

4 Results

4.1 Implementation

We subsampled the Northern Boreal Forest map without replacement to create our various samples. In order to get more easily interpretable parameter estimates, we centered and rescaled the locations of each observation. Originally the location units were meters-North and meters-East, but they were rescaled to 10km units. We also ran some models with log-transformed values to make the observations more closely normally distributed, but the log-transformation offered little-to-no advantage in prediction error or timing.

While running **spBayes**, we reviewed trace plots to ensure adequate convergence and mixing. The trace plots were generally pretty good. (See Appendix A.1 for a typical example.) We fit different models by varying starting values, trend parameters, sample size (n), and the number of knots (m). Most of the models resulted in similar prediction maps. The residual sum of squared error of the predicted map tended to improve in models with more knots, although more knots generally meant longer computing time.



Minimum	Maximum	Median	Mean	Std. Deviation
29.24	192.2	65.92	71.08	29.33

Figure 6: The first plot shows model's predicted values for 189 locations, and the histogram shows the density of the predicted values.

4.2 Maps

Since most models we ran created similar prediction maps, we present one as an example. This model used a sample size of $n = 10{,}000$ observations with $m = 265$ knots, took 2 days to complete 100,000 iterations, and had a residual sum of square error of just over 10,000. It was fitted using intercept, Northing, Easting and an interaction between Northing and Easting for main effects. The plots in Figure 6 show the resulting predicted values on the Northern Boreal Forest as well as their distribution. The model does tend to smooth some extreme values as we can see in the statistics, but overall, the predicted map is

quite similar to the sample map in Figure 5, and it compares favorably to the graphs and statistics in Figure 4.

Figures 7 and 8 illustrate ways to compare the prediction results of the model with the sample. Figure 7 shows the sample values; the contour lines were created using the predicted values. The prediction map captures most of the large scale trends, but it misses some of the smaller-scale patterns such as the higher precipitation clusters near (0, 0) on the centered scale. This is likely because the knots are spaced too far apart to capture that type of spatial clustering. Figure 8 shows the residuals at each of the prediction points. Red residuals indicate that the model predicts more precipitation than the original data, and blue residuals show where the model underestimated precipitation. Although there are a few dark red and dark blue residuals, there is no overall trend within the residual plot which implies that the model is not systematically missing some component of variability.

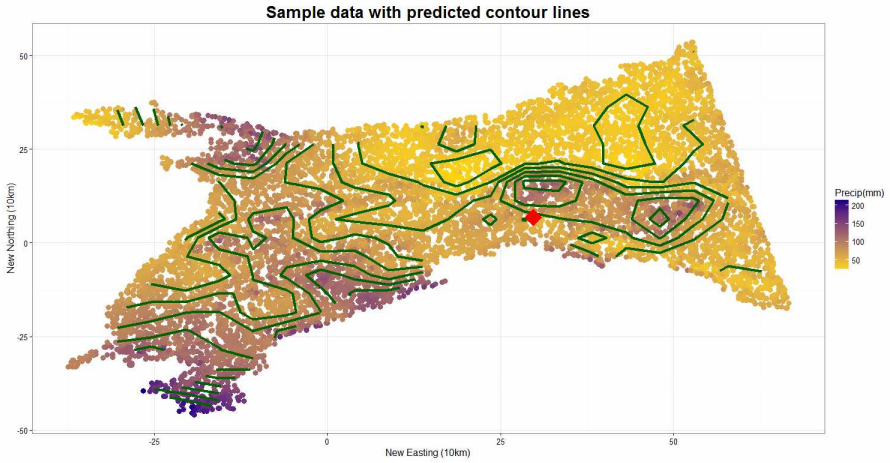


Figure 7: Map of subsample data with predicted contour lines from the predictive process model.

Figure 8: Map of residuals at 189 prediction locations. RSS = 10,959.64

4.3 Results of changing n and m

Figure 9 shows the results from various model runs, and the table in Appendix A.3 shows the change in time per MCMC iteration as a result of varying the sample size and number of knots. In general, for a fixed sample size, time per iteration increases as the number of knots increases. Figure 9 and Table A.3 also show the change in time per iteration increasing as sample size increases. However, increasing the number of knots increases the time per iteration exponentially while increasing the sample size increases the time per iteration linearly. The table in Figure 9 also shows that the residual sum of squares is most affected by an increase in the number of knots rather than an increase in sample size.

Table of varying model runs

Sample size: (n)	Knots: (m)	Iterations	Time: (s)	Seconds per iteration	Prediction Knots	Prediction time (s) per iteration	RSS
20,000	170	30,000	47,876	1.596	189	2.821	14,248.54
10,000	170	20,000	16,890	0.845	189	1.387	15,013.86
10,000	170	20,000	25,918	1.296	189	1.287	14,413.41
10,000	170	20,000	16,897	0.845	189	1.396	14,502.26
10,000	170	40,000	32,161	0.804	189	1.400	14,486.62
10,000	170	60,000	48,991	0.817	189	1.344	14,840.30
8,000	170	30,000	20,837	0.695	189	1.104	14,353.53
5,000	170	30,000	12,683	0.423	189	0.660	15,109.34
5,000	170	20,000	6,884	0.344	189	0.724	14,171.46
2,500	170	30,000	6,345	0.212	189	0.335	15,887.02
15,000	265	30,000	84,076	2.803	189	3.795	10,959.64
10,000	265	100,000	177,555	1.776	189	3.464	10,171.74
10,000	265	30,000	63,927	2.131	189	2.451	11,094.05
5,000	265	30,000	25,243	0.841	189	1.270	10,731.15
2,000	265	30,000	9,875	0.329	189	0.475	15,335.17
1,000	265	30,000	5,213	0.174	189	0.230	17,208.09
15,000	285	20,000	64,836	3.242	189	4.151	9,453.93
10,000	285	20,000	43,142	2.157	189	2.902	9,295.49
10,000	285	20,000	43,779	2.189	189	2.679	9,477.09
5,000	285	20,000	20,614	1.031	189	1.472	9,556.49
2,500	285	20,000	5,699	0.285	189	0.819	11,045.03
500	285	30,000	2,706	0.090	189	0.135	17,971.05

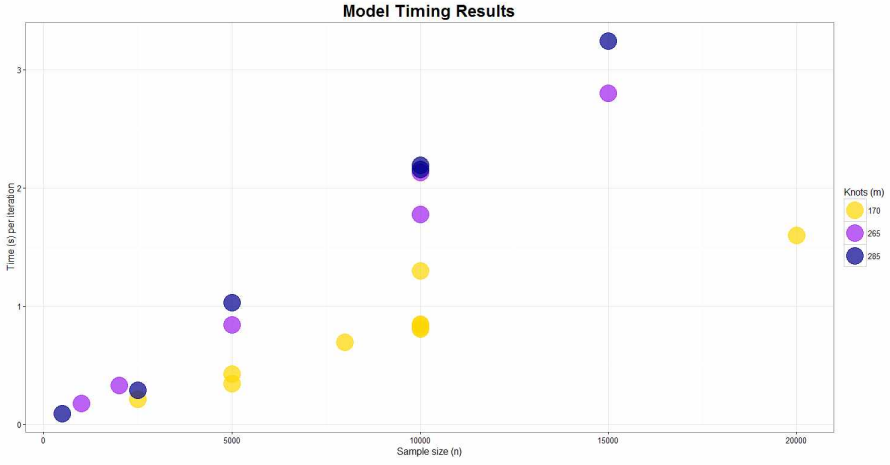


Figure 9: Table and scatterplot of model runs with varying sample sizes and number of knots.

5 Discussion

We used a Dell desktop computer running Windows 7 with an Intel Core i5 processor with 4 cores running at 3.20GHz and R version 3.2.2 for all models. Even with the dimension reduction achieved through the predictive process model, the models of the Northern Boreal Forest sometimes took days for the posterior distributions to fill in, which is why the trace plots in Appendix A.1 have small spikes around 20,000 iterations. Given enough time, the spikes seen in the trace plots would smooth out and the bimodal densities would likely become unimodal.

During MCMC iterations, the parameters are highly correlated which results in low acceptance rates as the proposals attempt to explore the parameter space. This makes it difficult to know whether the chains have converged or if they are stuck near a local maximum. Thus, long chains are needed because the β samples interact with ϕ and σ^2 which also interact with each other. The effects of the interactions can be seen in the model diagnostic plots in Appendix A.1.

In Appendix A.2, we have included results from an intercept-only model. This model used a sample size of $n = 15{,}000$ with $m = 265$ knots, took 23 hours to produce 30,000 iterations, and had a residual sum of squares of less than 11,000. The results are interesting in this case because the intercept-only model performs nearly as well as the models in which we included trend terms. Models that included Northing and Easting trend parameters produced credible intervals for the coefficients that did not include 0, so it seems that predictive process models are robust against mild model misspecification, particularly in the main effects process if predictions or maps are the primary goal. The tradeoff occurs in the covariance parameter estimates. If the model does not include trends, then the covariance function attempts to explain the relationship between observations over greater distances and the scale parameter, ϕ, in particular is affected.

5.1 Future work

The natural continuation of this paper is to fit additional models. Models that include quadratic trends in the main effects have yet to be run using this dataset, and there are more covariance functions besides the exponential function within **spBayes** that could be used. With a larger selection of models we could perform model selection using the deviance information criterion.

Additionally, applying the predictive process model to the full SNAP map of August 2013 precipitation and other maps seems like a good next step. Then applying the model to messier datasets instead of ones created by a model would be a good test of model performance.

References

Banerjee, Sudipto, Bradley P. Carlin, and Alan E. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press. ISBN: 9780203487808.

Banerjee, Sudipto and Montserrat Fuentes (2012). "Bayesian Modeling for Large Spatial Datasets". In: *Wiley Interdisciplinary Reviews. Computational Statistics* 4, pp. 59–66.

Banerjee, Sudipto et al. (2008). "Gaussian predictive process models for large spatial data sets". In: *Journal of the Royal Statistical Society*. B 70, pp. 825–848.

Cressie, Noel A. C. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley. ISBN: 9780471002550.

Daly, Christopher and Kirk Bryant (2013). *The PRISM Climate and Weather System. An Introduction*. Online document. URL: `http://www.prism.oregonstate.edu/documents/PRISM_history_jun2013.pdf`.

Daly, Christopher et al. (2008). "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States". In: *International Journal of Climatology* 28(15), pp. 2031–2064. ISSN: 1097-0088. DOI: 10.1002/joc.1688. URL: http://dx.doi.org/10.1002/joc.1688.

Eidsvik, Jo et al. (2012). "Approximate Bayesian inference for large spatial datasets using predictive process models". In: *Computational Statistics and Data Analysis* 56.

Finley, Andrew O., Sudipto Banerjee, and Alan E. Gelfand (2015). "spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models". In: *Journal of Statistical Software* 63 (13). URL: http://www.jstatsoft.org/.

Finley, Andrew O. et al. "Bayesian multivariate process modeling for prediction of forest attributes". In: *Journal of Agricultural, Biological, and Environmental Statistics* 13(1), pp. 60–83. ISSN: 1537-2693. DOI: 10.1198/108571108X273160. URL: http://dx.doi.org/10.1198/108571108X273160.

Higdon, Dave (2002). "Space and space-time modeling using process convolutions". In: *Quantitative methods for current environmental issues*, pp. 37–56.

SNAP (2015). *About SNAP*. International Arctic Research Center. URL: http://www.snap.uaf.edu/about/about (visited on 03/21/2016).

Stein, Michael L. (2008). "A modeling approach for large spatial datasets". English. In: *Journal of the Korean Statistical Society* 37(1), pp. 3–10. DOI: 10.1016/j.jkss.2007.09.001.

Stein, Michael L., Zhiyi Chi, and Leah J. Welty (2004). "Approximating likelihoods for large spatial data sets". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(2), pp. 275–296. ISSN: 1467-9868. DOI: 10.1046/j.1369-7412.2003.05512.x. URL: http://dx.doi.org/10.1046/j.1369-7412.2003.05512.x.

A Appendix

A.1 Model Diagnostic Plots

The following figures show trace plots and plots of parameters that were used to determine whether the models had converged and sufficiently mixed.

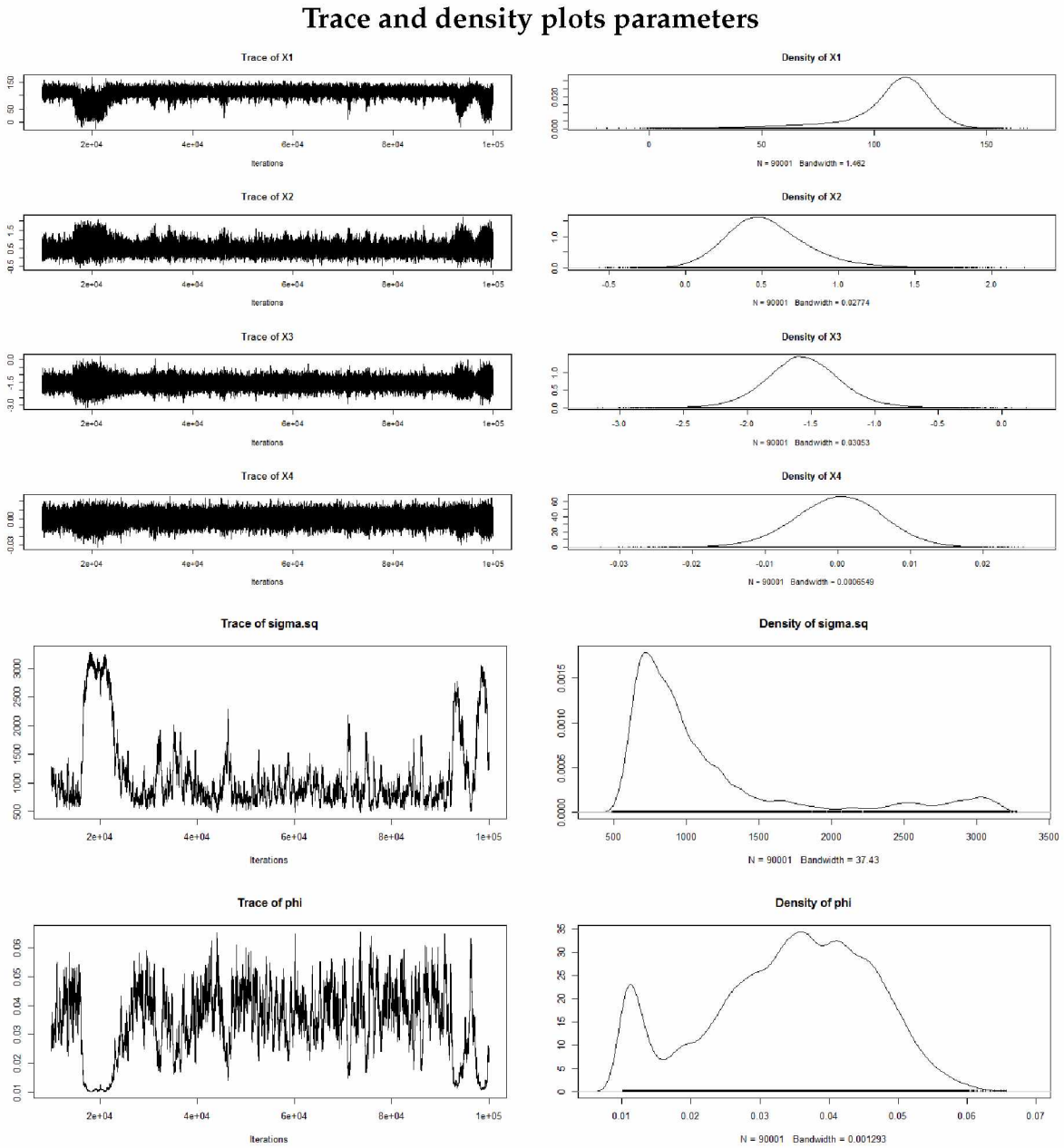


Figure 10: X1 is the intercept, X2 is Easting, X3 is Northing and X4 is the interaction.

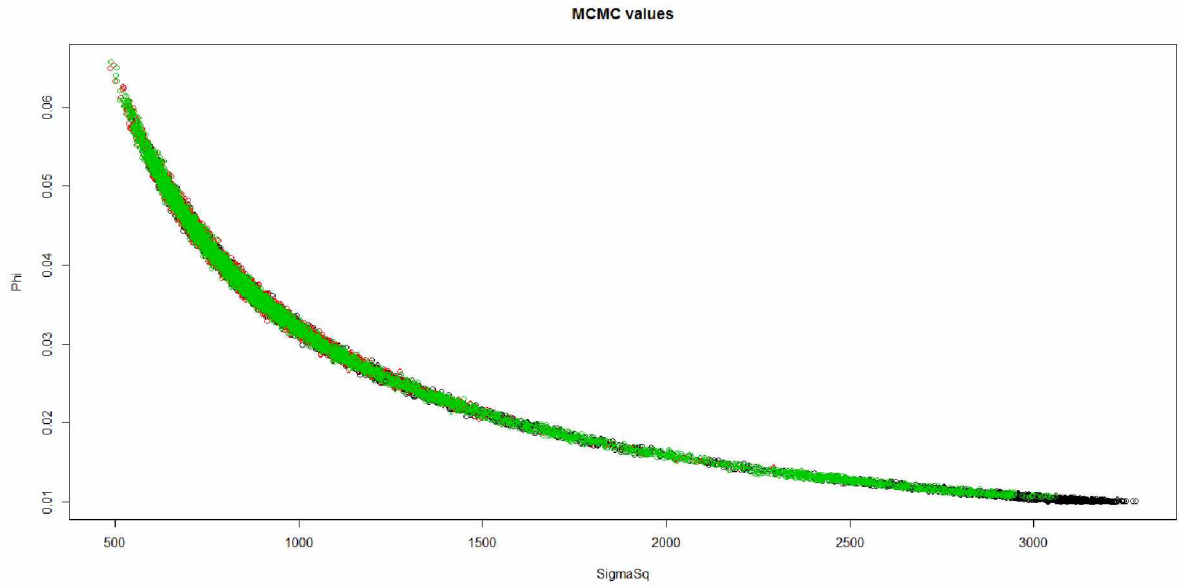
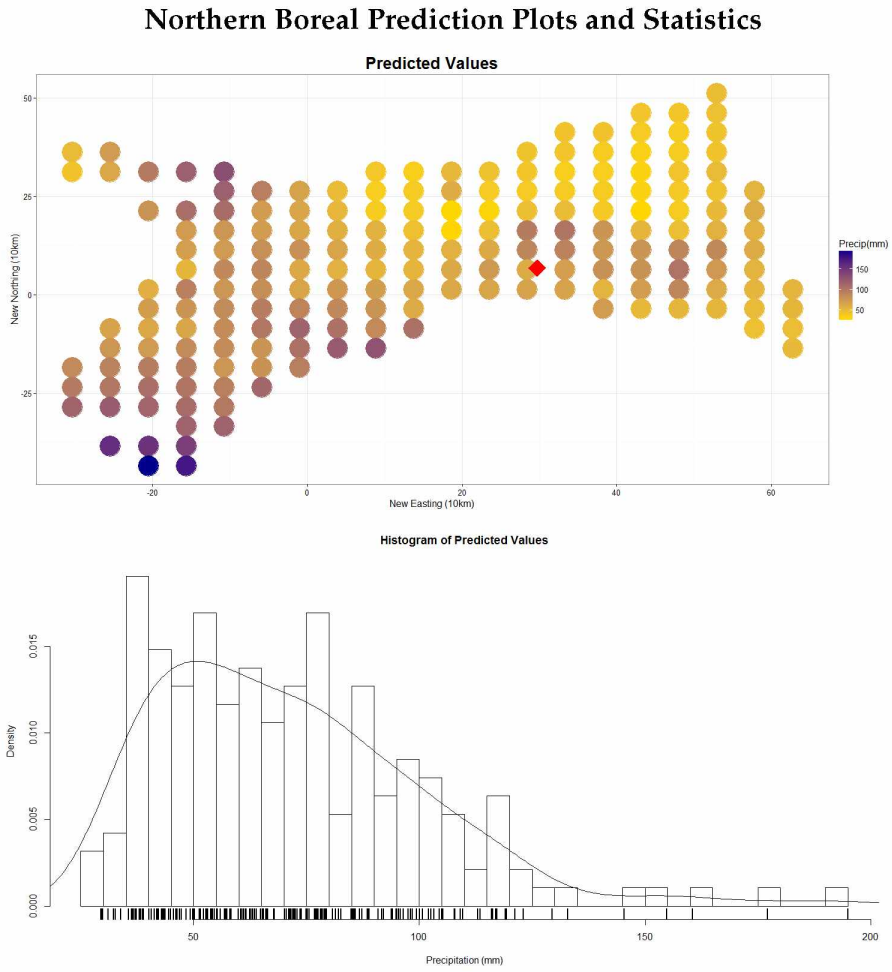


Figure 11: Values of ϕ and σ^2 at each iteration. The first third of the iterations are black, the middle third are red, and the final third are green. Note the strong curvilinear association.

A.2 No Trend Model Results



Minimum	Maximum	Median	Mean	Std. Deviation
29.47	195	66	71.04	29.27

Figure 12: These plots come from an intercept-only model fit to a subsample of $n = 15{,}000$ observations with $m = 265$ knots. The first plot shows model's predicted values at 189 locations, and the histogram shows the density of the predicted values.

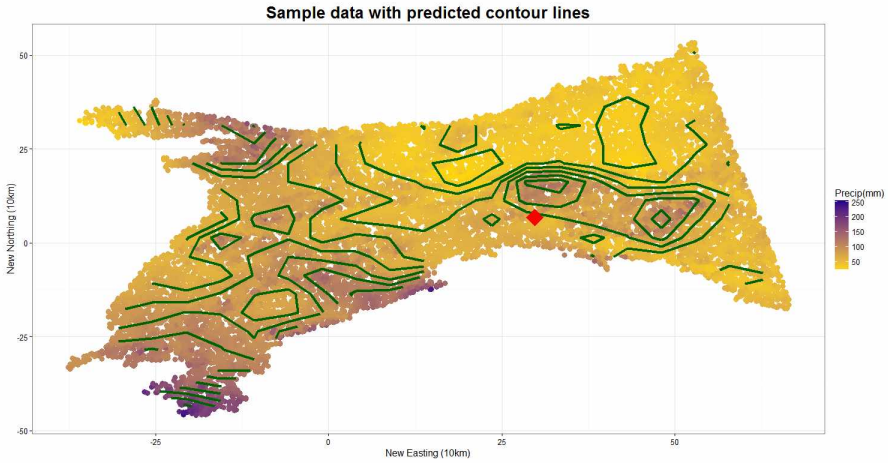


Figure 13: Map of subsample data ($n = 15{,}000$) with predicted contour lines from the predictive process model ($m = 265$) and RSS = 10,959.64

Figure 14: Map of residuals at 189 prediction locations from the $n = 15{,}000$ and $m = 265$ intercept-only model with RSS = 10,959.64

A.3 Table of varying model runs

Sample size (n)	Knots (m)	Time per iteration	Pred. knots	Pred. time per iteration
5,000	170	0.374	189	0.675
5,000	170	0.369	189	0.673
5,000	170	0.414	189	0.724
5,000	170	0.424	189	0.696
5,000	170	0.340	189	0.586
5,000	189	0.399	189	0.762
5,000	189	0.445	189	0.741
5,000	189	0.515	189	0.929
5,000	189	0.511	189	0.948
5,000	189	0.486	189	0.740
5,000	219	0.579	189	0.921
5,000	219	0.584	189	0.921
5,000	219	0.931	189	1.185
5,000	219	0.722	189	1.069
5,000	219	0.596	189	1.015
5,000	250	0.960	189	1.505
5,000	250	0.967	189	1.304
5,000	250	0.859	189	1.365
5,000	250	0.907	189	1.445
5,000	250	0.928	189	1.372
5,000	285	1.174	189	1.631
5,000	285	1.198	189	1.640
5,000	285	1.519	189	1.461
5,000	285	1.244	189	1.450
5,000	285	1.182	189	1.627
8,000	170	0.687	189	1.099
8,000	170	0.742	189	1.273
8,000	170	0.719	189	1.285
8,000	170	0.669	189	1.195
8,000	170	0.716	189	1.234
8,000	189	0.842	189	1.468
8,000	189	0.908	189	1.459
8,000	189	0.850	189	1.464
8,000	219	1.151	189	1.638
8,000	219	1.018	189	1.628
8,000	219	1.025	189	1.625
8,000	250	1.276	189	1.962
8,000	250	1.341	189	2.044
8,000	250	1.477	189	2.269
8,000	285	1.905	189	2.438
8,000	285	1.942	189	2.676
8,000	285	1.971	189	2.282